
Science Working Group on Data: A Data Distribution Workshop

*Graham Bothwell, Chairman of the EOS Science Working Group on Data
(graham.bothwell@jpl.nasa.gov)*

The EOS Science Working Group on Data (SWGd) arose from an initiative at the March 2000 meeting of the Science Working Group for the AM Platform (SWAMP) for an on-going evaluation of how well the current EOS data system can support the Terra, Aqua, and Aura missions for which it was designed. Information about the SWGD, including reports of workshops, can be found on the recently created SWGD web site, at <http://swgd.gsfc.nasa.gov>.

As reported in the November/December 2000 issue of *The Earth Observer*, the inaugural SWGD workshop was held on June 01-02, 2000, at the Goddard Space Flight Center (GSFC). Although that event concentrated primarily on data processing requirements for NASA's Terra mission, it also concluded that innovative approaches would be needed to meet data distribution needs. It was recommended "that there be a meeting about six months from now to address data distribution status and archive access needs." This meeting took the form of a Data Distribution Workshop, which was held on February 1, 2001, at Fort Lauderdale, Florida, in conjunction with the EOS Investigators Working Group (IWG) meeting.

The workshop was designed to assess the current status of Terra data distribution and identify immediate and foreseeable obstacles to meeting user data needs, and to identify critical needs and areas for improvement and approaches for new development.

Stakeholder status and feedback

The first part of the workshop was a gathering of status and feedback from all the parties participating in EOS data distribution, including NASA Headquarters; the EOS Program Office and ESDIS Project at GSFC; the DAAC User Working Groups (UWGs); and DAAC management.

All of the EOS DAACs have User Working Groups (UWGs). Besides facilitating the requirements and issues of the user community, the UWGs have a vital role assisting the DAACs to determine user needs, interface requirements, and priorities. Reports from the UWGs to the workshop indicated that the ordering system works well (within its inherent limitations), that data is easy to select, and that user services support is generally good. There are also numerous issues still being worked or remaining to be worked, such as, to list but a few, difficulties in ordering collocated data from multiple sensors; concern about limits to the volume of distributed products; and the lack of similar file naming conventions between instruments.

The immediate data distribution needs at the DAAC-based distribution systems are highly individual. While distribution capabilities at the JPL DAAC are adequate, the larger DAACs at GSFC, NASA Langley, and the EROS Data Center (EDC) have not yet fully achieved their potential due to a range of issues, most of which arise from the newness of the Terra mission. The more critical issues are in the process of being addressed, such as problems with high-volume orders. There are also long-term capacity issues.

Selected issues and recommendations

The second part of the workshop involved discussing the process towards resolving the various issues in data distribution, and making specific recommendations. To facilitate this, the meeting divided into three discussion groups to address selected priority issues.

A. Software tools to facilitate distribution and early use of data

The EOS data products are not always regarded as easy to read and handle, and current tools to assist with this are limited in capability. The range of tool requirements includes data product search and order tools; format conversion tools; and data manipulation tools e.g. reprojection. A three-step process was recommended for providing tools that offer essential basic capabilities, namely (1) a survey of available tools, including those available from commercial software packages; (2) a commitment by the DAACs to support selected tools; and (3) development of new tools where capabilities do not already exist.

The group concluded that, because the provision and support of software tools will fall ultimately to the DAACs, responsibility for tool definition and advocacy should reside with them, in conjunction with their UWGs, with funding for tool development by the community sought through budget augmentation.

B. User modeling

The primary application of user modeling is in prioritizing and making resource allocations at the DAACs more effective, such as to aid in identifying current and future stress points; to develop mitigation strategies; and to help users and DAACs become more efficient. Without a successful modeling technique, systems can be wrongly sized and funding mistakes can readily occur. There is currently no model used to forecast system evolution. Steps to overcome this include (1) better use of existing available metrics; (2) development of a suitable range of alternative modeling techniques, including the extensive model developed by and described to the workshop by Bruce Barkstrom; (3) discussion based on presentations to the various UWGs; and (4) development of profiles for different kinds of users.

C. Creative solutions to current and projected distribution obstacles

This topic was designed to suggest techniques for resolving issues in data distribution not addressed by current plans or developments. The different classes of solution that could be investigated include:

- Contributions from groups other than the DAACs might include value added products; helping to distribute standard products, e.g. partial or full mirror sites.
- Direct broadcasting might be facilitated more, along with software tools to use the data received, and the data recipients encouraged to redistribute their data.
- Greater distribution efficiency may be achieved through new technological mechanisms such as data pools, data mining, coincident data searches, and various other ways.
- Not-for-profit organizations may be able to assist with creative financing.
- Braking mechanisms may be useful, such as a sliding scales of data availability depending on the size of data sets ordered.
- Certain products might be developed in more innovative manners.
- Improved data user tools were discussed above.
- Anticipation of needs through better user models was discussed above.

Along with these ideas were suggestions for encouraging creative solutions:

- It was felt that the DAACs should take the lead in better defining existing and future needs of the users, using the UWGs and other mechanisms.

-
- A full spectrum of organizations and groups needs to be involved in continuing the momentum of ideas initiated by the workshop.
 - There needs to be a mechanism to ensure a continued supply of data sets from non-traditional suppliers.
 - A programmatic mechanism should decide what resources are needed to maintain the data holdings and ensure long term archiving.
 - It is important to allow for innovative solutions, including unconventional ideas.
 - The capabilities of the non-NASA community can be engaged where relevant.

In many instances, the realization of creative solutions will be part of the evolution of NewDISS, the system that is planned to handle NASA's Earth Science data processing in the future. It is a concept for a distributed, flexible, responsive system that allows for a spectrum of heterogeneous approaches, utilizing key standard interfaces to facilitate a workable across-the-board infrastructure. NewDISS is still in the formulation stage, but evolution towards it is already emerging.

Conclusion and next steps

The success of the SWGD depends upon a genuine interaction between the instrument and science teams represented by the SWGD and the program and project management, so that issues relating to success of the respective EOS missions are resolved effectively. The prospects for this are promising, based on the senior level of participation in the workshop by representatives from NASA Headquarters and the GSFC Program and Project Offices.

The core of future SWGD activities is embodied especially in the above-reported discussion on creative solutions. The discussions at the current workshop represent only a first step, and are primarily at the level of potential possibilities that will require ongoing elaboration and review to ensure a successful evolution of progress. A future meeting on creative solutions is therefore recommended, possibly in conjunction with the next IWG meeting.

There is also scope to continue the exploration of topics at future workshops, typical topics including EOS user models, software tools, and Terra archiving.

In conclusion, the SWGD seeks to work with the existing organizational, developmental, and operational structures to assist the EOS missions. In doing this, it is necessary to work for a community consensus and to assist in communicating that consensus and its related proposals. It is important that no opportunities be left unutilized or underutilized. This workshop made a good start to addressing the many issues associated with data distribution.